



Optical-Flow Based Nonlinear Weighted Prediction for SDR and Backward Compatible HDR Video Coding

David Gommelet, Julien Le Tanou, Aline Roumy, Michael Ropert, Christine Guillemot

► To cite this version:

David Gommelet, Julien Le Tanou, Aline Roumy, Michael Ropert, Christine Guillemot. Optical-Flow Based Nonlinear Weighted Prediction for SDR and Backward Compatible HDR Video Coding. IEEE Transactions on Image Processing, In press, pp.1-16. 10.1109/TIP.2019.2945685 . hal-02378053

HAL Id: hal-02378053

<https://hal.science/hal-02378053>

Submitted on 24 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optical-Flow Based Nonlinear Weighted Prediction for SDR and Backward Compatible HDR Video Coding

David Gommelet, Julien Le Tanou, Aline Roumy, *Member, IEEE*, Michaël Ropert, and Christine Guillemot, *Fellow, IEEE*,

Abstract—Tone Mapping Operators (TMO) designed for videos can be classified into two categories. In a first approach, TMOs are temporal filtered to reduce temporal artifacts and provide a Standard Dynamic Range (SDR) content with improved temporal consistency. This however does not improve the SDR coding Rate Distortion (RD) performances. A second approach is to design the TMO with the goal of optimizing the SDR coding rate-distortion performances. This second category of methods may lead to SDR videos altering the artistic intent compared with the produced HDR content. In this paper, we combine the benefits of the two approaches by introducing new Weighted Prediction (WP) methods inside the HEVC SDR codec. As a first step, we demonstrate the interest of the WP methods compared to TMO optimized for RD performances. Then we present the newly introduced WP algorithm and WP modes. The WP algorithm consists in performing a global motion compensation between frames using an optical flow, and the new modes are based on non linear functions in contrast with the literature using only linear functions. The contribution of each novelty is studied independently and in a second time they are all put in competition to maximize the RD performances. Tests were made for HDR backward compatible compression but also for SDR compression only. In both cases, the proposed WP methods improve the RD performances while maintaining the SDR temporal coherency.

Index Terms—High dynamic range (HDR), Compression, Tone-Mapping (TMO), Backward-Compatible, HEVC, Weighted Prediction, Nonlinear Illumination Compensation

I. INTRODUCTION

HIGH Dynamic Range (HDR) imaging allows the reproduction of much brighter and darker lights, or luminance levels, than Standard Dynamic Range (SDR) imaging. The ultimate aim of HDR imaging is to capture and reproduce all the dynamic range of the visible light perceivable by the human eye. This dynamic range goes from 10^{-6} to 10^8 nits thus providing 14 orders of luminance magnitude. The human is only able to perceive all this dynamic through several eye adaptation mechanisms, and without these adaptations, it can instantaneously perceive 5 orders of luminance magnitude. Recently, several HDR displays were proposed, with a minimum of 5 orders of luminance magnitude, to at least match the instantaneous human eye vision. In contrast, SDR devices can only process 3 orders of luminance magnitude.

To compress HDR content, several solutions are available. The first solution, adopted by the MPEG standardization group, consists in applying a PQTF (perceptual quantizer transfer function) on the HDR values to obtain a new content represented on 10 bits. Then, this content is compressed with legacy SDR compression standard such as HEVC. This solution yields good compression performances and is compatible with new HDR displays. However, the 10-bit content cannot be visualized on SDR displays.

A second solution to address both HDR and SDR displays is to convert HDR content into a SDR signal using a Tone Mapping Operator (TMO) and transmit both the HDR and SDR in separate bitstreams. This solution called simulcast is clearly not optimal in terms of compression performances.

Finally, in a better alternative, the HDR input is first tone mapped into an SDR version which is encoded with a legacy codec such as HEVC and sent along with metadata containing TMO parameters. This signal can be decoded with legacy SDR displays or it can be inverse tone mapped using the metadata to yield an HDR reconstructed content. This approach is referred to as single layer backward-compatible scheme. In addition to the SDR content and the metadata, one can also send an enhancement layer containing the HDR residue to improve the reconstruction of the HDR content. Compression schemes using an enhancement layer are referred to as two layer (or scalable) schemes. We showed in a previous contribution [1], that a two layer scheme can be optimally separated into a TMO optimization in a single layer scheme followed by a rate allocation step.

In this paper, as in [2]–[5], we focus on the design of a single layer scheme for HDR video compression. Our goal is to design a scheme that would yield both high rate-distortion (RD) performances and high fidelity between the tone mapped SDR video and the original HDR video, for example by preserving temporal variations in the input HDR video sequence. As a matter of fact, many authors tried to improve the temporal coherency of tone mapped SDR videos by applying a temporal filtering [2]–[4], [6], [7], or more elaborate operators [8], [9] on the TMOs of each frame. Although these filtering operators ensure a better temporal consistency of the SDR frames, this has been shown to have little impact on the RD performance of the compression scheme. The authors in [5] instead aim at designing a TMO that improves the RD compression performance but does not preserve the illumination variations present in the HDR content.

D. Gommelet, A. Roumy and C. Guillemot are with Inria, Rennes-Bretagne-Atlantique, Rennes 35042, France (e-mail: firstname.lastname@inria.fr).

D. Gommelet, M. Ropert and J. Le Tanou are with Mediakind, Saint Jacques de la Lande 35136, France (e-mail: firstname.lastname@mediakind.com).

Manuscript received ...; revised ...

In this paper, we show that, to better preserve the artistic intent in the tone mapped SDR video, a better alternative consists in applying standard tone mapping operator and in using appropriate weighted prediction mechanisms for compensating for illumination variations in the SDR signals. This allows to improve the SDR RD performances and therefore the HDR ones, while preserving the artistic intent present in the HDR signal rather than altering the temporal effect.

Classical WP consists of computing a linear transformation between the source image and a reference image without motion compensation. The transformation can be global, i.e. applied on the whole frame, as implemented in the H264 and HEVC codec [10, Chap. 5]. It can also be local, either by storing many possible global WP frames [11]–[13] as multiple reference frames, or by computing one WP per block [14]. To avoid the overhead of sending the WP parameters for each block, the parameters can be computed on neighboring blocks [15]–[17]. This, however, increases the encoding cost of the residue due to the lack of accuracy of the WP parameters. Still, in all these works, the WP parameters are computed without taking into account any motion compensation and linear functions, for global WP or local WP are used.

In this paper, we propose both a new algorithm to compute the WP parameters and different WP models. Instead of computing the WP directly between the consecutive images, the proposed WP algorithm uses an optical-flow based motion compensation. Furthermore, to predict the complex temporal variations between SDR frames, two new WP models, implying syntax and decoder modifications, are introduced: a global non-linear WP and a local non-linear WP. The paper is organized as follows. Section II demonstrates the equivalence between the proposed global non-linear WP and the temporally constrained TMO [5]. An overview of the proposed WP algorithm is proposed in section III with further details on the HEVC implementation in section IV. Finally, experimental results are presented in section V.

II. TEMPORALLY CONSTRAINED TMO AND WEIGHTED PREDICTION

A. Temporally constrained TMO

In [5], the authors proposed to improve the TMO optimization of a single layer compression scheme described in [3] by accounting for the rate R of the tone mapped SDR video. They expressed this rate as a function of the temporal activity $C(S_t)$ of the SDR content, i.e. as:

$$R \propto C(S_t) = \sum_{i,j} \left(S_t(i,j) - M(S_{t-1}(i,j)) \right)^2 \quad (1)$$

with S_t the current tone mapped SDR frame, S_{t-1} the previous tone mapped SDR frame, (i,j) the pixel position in the respective frame, and M the motion compensation between S_{t-1} and S_t , which is computed using an optical flow [18]. This estimator is illumination-invariant. This requires introducing, in the optimization problem, additional variables, which model the illumination changes. A solution to this augmented problem is obtained by a primal-dual minimization approach, which iterates between the estimation of the illumination changes

and the estimation of the motion field [18]. When computing the current TMO, called F_t , the SDR frame S_t is unknown and the motion compensation cannot be computed. This is illustrated in Figure 1. To overcome the motion compensation

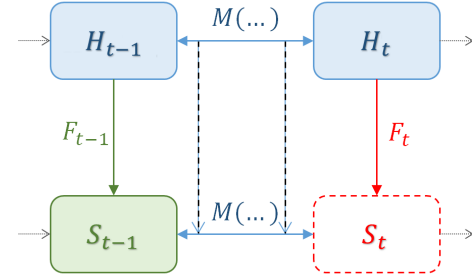


Fig. 1. Tone Mapping computation for two consecutive HDR frames as in [5]

problem, the authors estimated M between the HDR frames, H_{t-1} and H_t , and applied it to the SDR frame S_{t-1} . However, the constraint (1) still needs to be expressed as a function of the TMO F_t . To do so, they parameterize F_t as a piecewise linear tone curve, but also make a coarse assumption over the distribution of the HDR values in each bin of the HDR histogram. They assumed that all values inside one bin have the exact same value, which is the center of the bin.

Adding the rate constraint (1) to the distortion optimization problem proposed in [3], the optimization problem in [5] is:

$$\begin{aligned} \min_{s_k} \quad & \sum_{k=1}^N (p_k \cdot s_k^{-2}) + \lambda \cdot C(S_t) \\ \text{s.t.} \quad & \sum_{k=1}^N s_k = \frac{2^{nb} - 1}{\delta}, \quad \forall k, s_k > 0 \end{aligned} \quad (2)$$

with nb the bit-depth for the SDR frame, δ the selected interval for the HDR frame histogram, p_k the probability of the k -th histogram bin, s_k the TMO (F_t in Fig 1) slope for the k -th bin and λ a Lagrangian multiplier to simplify the constrained problem into an unconstrained one. The remaining constraint ensures that the TMO curve covers all the SDR dynamic.

Focusing only on the minimization of the rate constraint (1) (i.e. the specific case where $\lambda = \infty$ in (2)), and ignoring the simplifications made in [5], then (1) becomes:

$$C(S_t) = \sum_{i,j} \left(S_t - M(S_{t-1}) \right)^2 \quad (3)$$

Note that, in this equation and the following ones, we removed the pixel index (i,j) for a simpler notation.

$$C(S_t) = \sum_{i,j} \left(F_t(H_t) - M(F_{t-1}(H_{t-1})) \right)^2 \quad (4)$$

If one defines BV the brightness variations from $M(H_{t-1})$ the motion compensated HDR frame H_{t-1} , to H_t , then $H_t =$

$BV(M(H_{t-1}))$ and:

$$C(S_t) = \sum_{i,j} \left(F_t(H_t) - M(F_{t-1}(M^{-1}(BV^{-1}(H_t)))) \right)^2 \quad (5)$$

$$= \sum_{i,j} \left(F_t(H_t) - F_{t-1}(BV^{-1}(H_t)) \right)^2 \quad (6)$$

where (6) follows from the fact that M^{-1} can be permuted with F_{t-1} since M^{-1} is a pixels displacement and F_{t-1} a global function over the pixel values. Thus, the optimal solution becomes obvious:

$$F_t(\dots) = F_{t-1}(BV^{-1}(\dots)) \quad (7)$$

The TMO F_{t-1} is a global invertible function and BV^{-1} usually is a pixelwise non-invertible function. The challenge here, is to reproduce BV^{-1} with the global invertible function F_t .

Optimizing (2) with $\lambda = \infty$ is clearly not optimal in the context of backward compatible compression. First, the term considering the HDR distortion is then completely ignored which is clearly not optimal for RD performances. Secondly, the visual quality of the resulting SDR video would be strongly impacted. Using the optimal solution, all global luminance variations over an HDR video would disappear in the SDR video. It is therefore necessary to compromise between distortion and rate minimization but also temporal regularization. In practice the authors in [5] empirically chose $\lambda = 0.1$ for their experiments.

B. Weighted Prediction vs. Temporally constrained TMO

The aim of the proposed weighted prediction is to improve RD performances of the temporally constrained TMO proposed in [5] but with an additional constraint on the SDR temporal quality or consistency. The proposed method shares many similarities with this TMO but also differs on key points, as illustrated in Figure 2. Indeed, using a WP avoids

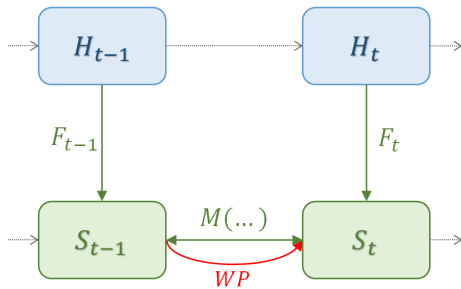


Fig. 2. Weighted prediction computation for two consecutive HDR frames as proposed

to compromise between distortion minimization and temporal regularization. It turns the constrained problem (2) into two consecutive problems, first minimizing the HDR distortion of H_t using the TMO [3] and in a second step, inside the video codec, minimizing the brightness variations between S_{t-1} and S_t with a WP. Consequently, the proposed strategy is easily adaptable to any TMO, contrary to [5]. One can use a TMO that preserves the SDR perceptual quality or a TMO that

minimizes the distortion under a rate constraint for the still image and thus benefit from a reduction of spatial and temporal complexity. Having two steps also allows to know the two SDR frames before computing the WP and therefore allows to directly compute the optical flow between S_{t-1} and S_t . One could argue that the computation of the optical flow is more accurate on the HDR level yet, experimentally, we didn't found noticeable differences between the computed motions.

Another advantage of the proposed WP is to preserve the temporal consistency of the SDR content. Indeed, in [5] the frame S_t is deliberately modified for compression purposes, therefore impacting the original temporal consistency when compared to HDR content. Using a weighted prediction, the frame S_{t-1} is used to predict S_t directly inside the encoding loop. Therefore, the original input SDR video is preserved.

Focusing on the second step, minimizing the brightness variations with a WP is really similar to minimizing (3):

$$C(S_t) = \sum_{i,j} \left(S_t - WP(M(S_{t-1})) \right)^2 \quad (8)$$

$$= \sum_{i,j} \left(S_t - WP(M(F_{t-1}(H_{t-1}))) \right)^2 \quad (9)$$

$$= \sum_{i,j} \left(S_t - WP(M(F_{t-1}(M^{-1}(BV^{-1}(H_t))))) \right)^2 \quad (10)$$

$$= \sum_{i,j} \left(S_t - WP(F_{t-1}(BV^{-1}(H_t))) \right)^2 \quad (11)$$

$$= \sum_{i,j} \left(S_t - WP(F_{t-1}(BV^{-1}(F_t^{-1}(S_t)))) \right)^2 \quad (12)$$

And therefore, the goal is to find WP such that:

$$WP(F_{t-1}(BV^{-1}(F_t^{-1}(x)))) = id_X \quad (13)$$

with id_X the identity map.

Comparing (7) and (13), one realizes that the proposed WP needs to reproduce a composition of three functions instead of two for (7). As explained previously, the most difficult part is to estimate the pixelwise non-invertible function BV since F_{t-1} and F_t are non-linear invertible functions. Approximating WP by a linear function, as currently done in HEVC, seems suboptimal as it is a composition of three functions, each of them being non linear. Therefore, our first proposal is to use a non-linear function as WP model in HEVC. To the best of our knowledge, global non-linear WP has never been proposed so far. This WP is supposed to improve the RD performances of [5] since the compromise between HDR distortion and SDR temporal regularization is removed. In addition, the proposed non-linear WP is not limited to invertible functions since the HDR content reconstruction is not needed.

Regarding the second proposed WP model, to better estimate the function BV , we extend our global approach to non-linear local WP. As explained previously, global and local linear WP have already been studied however, for comparison purposes, we also implemented these solutions in our framework. In this way, we can better assess the benefits of non-linear WPs.

III. NEW MODELS OF WEIGHTED PREDICTION

We now describe the new WP algorithm and the new WP modes. Here, WP modes refer to the signalization of the WP function and its parameter, whereas the WP algorithm is the method used to estimate the WP model parameters. Note that, the proposed WP algorithm only implies encoder modifications and is therefore compatible with the current HEVC standard. In contrast, the new WP modes require decoder modifications and thus, a modification of the HEVC standard.

Figure 3 summarizes the WP algorithm and WP modes presented in the following sections. The HEVC standard algorithm is improved with an optical flow based algorithm and the WP modes are extended from one type to four.

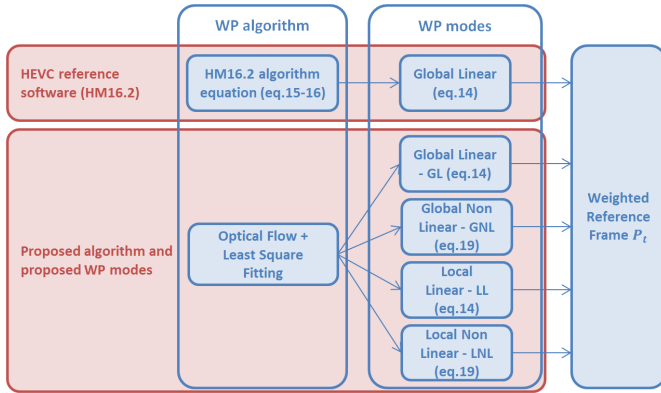


Fig. 3. Weighted Prediction Scheme - WP algorithms and WP modes.

A. HEVC Weighted prediction

In the HEVC reference software (HM16.2), the WP is applied to the frame used in the reference list. This WP is a global linear function directly derived for the entire reference frame S_{t-1} by using the average and variance of the luminance values of the frame to be predicted S_t and the reference frame S_{t-1} . Therefore, these values do not account for motion compensation between the two frames. The weighted reference frame P_t for S_t is obtained with:

$$P_t(i, j) = W \cdot S_{t-1}(i, j) + O \quad (14)$$

as reported in [10, eq (5.9)] and with (i, j) the pixel position in the frame. The frame P_t is later used for motion compensation. Then, WP is applied if it allows to reduce the sum of absolute differences (SAD) by at least 1% i.e. if

$$\frac{SAD(S_t, P_t)}{SAD(S_t, S_{t-1})} \leq 0.99 \quad (15)$$

This method is simple and fast but provides an imprecise WP and thus an imprecise luminance compensation.

B. Proposed Weighted Prediction

The proposed WP shares similarities with the WP implemented in the reference Software (see Section III-A). First, it is also applied on the previous frame S_{t-1} and second, it satisfies (14). However, there are several ways in which the proposed

WP differs from the classical one. First, the WP parameters W, O may be local. Second, the WP parameter computation takes into account motion compensation since it depends on the motion compensated reference frame $M(S_{t-1})$, and not the reference frame directly S_{t-1} . For fair comparison with [5], the proposed WP parameters is computed by using the same optical-flow M [18].

As in the HM16.2, the WP activation in the proposed encoding process is determined by a threshold:

$$\frac{SAD(S_t, M(P_t))}{SAD(S_t, M(S_{t-1}))} \leq 0.95 \quad (16)$$

As opposed to (15), the SADs are computed between S_t and the motion compensated frame S_{t-1} , thus providing much lower values. In practice, a threshold of 0.99 was leading to an over-activation of the WP with (16). For these reasons, we decrease the threshold to 0.95. The same activation threshold is used for all the following WP modes.

1) *Global and Local WP*: Using the same WP for the entire frame results in a small rate overhead cost since we send only few coefficients by frame. However, using a global WP suggests that the brightness variations between two frames is global. For many contents, this assumption is false, hence the use of local WP. As explained in the introduction, the local WP can be implemented with different methods. In this case, we choose to send the local WP as metadata within each CTU (Coding tree unit/maximum block size) in HEVC.

We implemented 4 WP modes: global linear (GL), global non-linear (GNL), local linear (LL), local non-linear (LNL). GL and LL WPs have already been proposed in the literature but without motion compensation. So, we implemented GL and LL WPs to assess the RD performance gains due to luminance compensation and also to compare their RD performances to the ones obtained with the GNL and LNL WPs.

The next sections detail the derivations of linear and non-linear WP parameters. Global and local methods are computed using the same process, only the input pixels change. In the global case, all matching pixels between the motion-compensated frame $M(S_{t-1})$ and the current frame S_t are plotted on a 2D graph thus giving a point cloud. In the local case, the process is exactly the same, except that we plot the matching pixels of a 64x64 block instead of the entire frame.

2) *Linear WP*: For the linear WP modes, the weighted reference frame is computed using (14). To derive the parameters W and O , all matching pixels of the motion-compensated frame and the current frame (64x64 block in the local case) are plotted on a 2D graph, as illustrated in Figure 4. Then W and O are obtained with linear least square fitting of this cloud point. In practice, we chose to constrain the minimum and maximum values of W and O :

$$-128 \leq W \leq 127 \quad (17)$$

$$-256 \leq O \leq 255 \quad (18)$$

Figure 4 also shows some examples of GL and LL WPs. Some samples of the corresponding SDR videos are visible in Figure 5. For the GL WP, the metadata for each frame only consists in the two coefficients W and O . On the other hand,

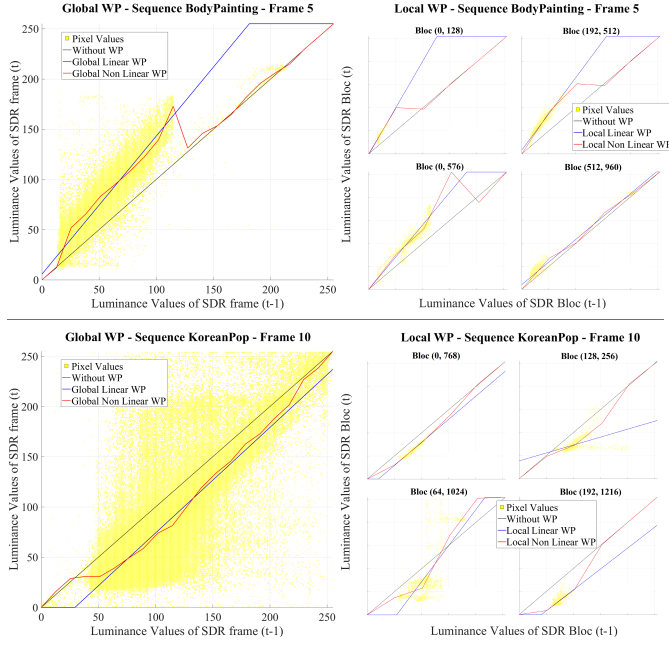


Fig. 4. Computed WP for two images (top: BodyPainting, bottom: KoreanPop). The left charts show the global linear and global non-linear WP functions (using 20 slopes) for the entire frame. The right charts show the local linear and local non-linear WP functions (using 5 slopes) for four different 64x64 blocks

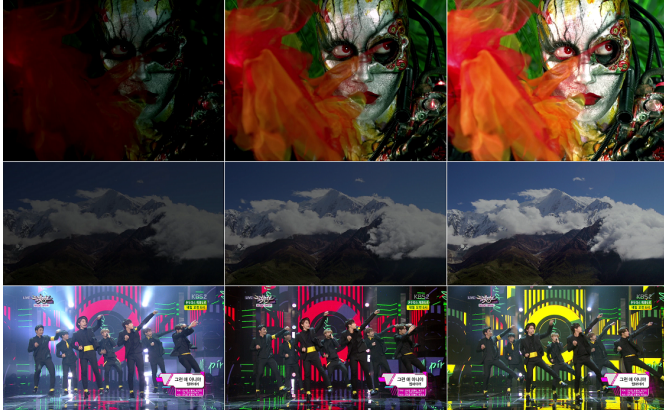


Fig. 5. Samples of tested SDR sequences. Each row represents three selected frames of each sequence. From top to bottom: BodyPainting Frame 4-6-8, Tangerine Frame 1-10-20, KoreanPop Frame 1-4-18

the metadata for the LL WP depend on the video size but also on the activation frequency of the WP. The coefficients W and O are sent for each CTU where the WP improves the RD performances.

3) *Non-linear WP*: For the non-linear WP modes, as in [5], we chose to parameterize WP as a piecewise linear function. Therefore, the weighted reference frame P_t is computed using the following equation:

$$\forall i, j \cap \forall S_{t-1}(i, j) \in [\delta.k, \delta.(k+1)[$$

$$P_t(i, j) = (S_{t-1}(i, j) - \delta.k) \cdot s_k + \delta \cdot \sum_{j=0}^{k-1} s_j \quad (19)$$

with s_k the k -th slope in the piece wise linear function and δ the interval where each slope applies. In practice, WP always

spans all the possible values of S_{t-1} (0 to 255). Therefore, $\delta = \frac{255}{n}$ with n the chosen number of slopes.

As for the linear methods, all matching pixels of the motion-compensated frame and the current frame (64x64 block in the local case) are plotted on a 2D graph and the slopes s_k are obtained with piecewise linear least square fitting of the obtained cloud point. An example of GNL and LNL WP is visible in Figure 4. The value of n was fixed empirically for both methods, we chose $n = 20$ for the GNL WP and $n = 5$ for the LNL WP.

Knowing n , the WP function is fully determined with the slopes s_k . These slopes are represented with real values and therefore need many bits for their representation. In practice, we chose to represent the n slopes with $n+1$ ordinates. Indeed, the ordinates can be easily rounded without much precision loss. Therefore, using the GNL WP, the metadata consists in 19 ordinates. As the first and the last ones are always 0 and 255, they are ignored. As for the LL WP, the metadata of the LNL WP depend on the input size and efficiency of the WP. Each CTU where the WP improves the RD performances is sent with 4 ordinates.

IV. HEVC IMPLEMENTATION

The WP algorithm and WP modes are computed once, as a first step, on the uncompressed video using a Matlab implementation. Then, in a second time, all these WP modes are parsed and used in the HEVC reference software (HM16.2) to encode the video. This method allows to test multiple strategies in HEVC without recomputing all the WP modes parameters and optical flows. This section details the specificity of the HEVC implementation, especially the chosen signaling for each mode.

A. Frame Level Activation

The first strategy used consists in choosing the WP mode at a frame level, as illustrated in Figure 3. For the global methods (GL, GNL), the WP mode is chosen if the SAD criterion (16) is satisfied. Instead, for the local methods (LL or LNL), each frame meeting (16) is compressed twice, one time without WP, and a second time with the current WP (LL, or LNL). The frame with the best RD performances is then encoded along with its metadata.

For the linear WPs, the coefficients W and O use a real representation requiring many bits, in practice, we choose to round up these values. Empirically, we found that the most efficient way was to round them after a multiplication by 2^5 . The coefficient W then needs 13 bits and the coefficient O 14 bits. For the non-linear WPs, each ordinate value requires 8 bits. To reduce the overhead cost, first the difference of the ordinates is computed and the mean value (δ) is subtracted (see (19)). This value is then encoded with the VLC (Variable Length entropy Coding) in the HM16.2. This allows to compress the coefficients at a rate smaller than 8 bits per slope.

For the global WPs and local WPs, we just signal the metadata for each frame where the WP applies. Regarding the local WPs, we found that it was much efficient to signal

an one-bit flag for each CTU then sending the WP when it improves the CTU RD performances. Therefore, each CTU is compressed with and without the WP then encoded with the corresponding flag and potentially with the WP parameters.

As mentioned previously, the brightness between two frames varies greatly depending on the input, even in the same video. To further improve the RD performances, we also choose to put all these WP modes in competition since each of them outperforms the other ones for specific frames. To do so, each frame is compressed with each WP mode and the best one is encoded along with the frame. This competition mode needs an additional signaling at frame level, indeed an index is sent to signal the chosen WP mode. When using four competing WP modes the index uses two bits. We also put only two modes in competition, and thus with an one-bit index.

The algorithm 1 details the HEVC signaling for all WPs and for the competition modes when we choose the WP at a frame level.

Algorithm 1 Frame Level Activation - HEVC Signaling at the encoder side

```

for all Frame do
  Signal if the WP is activated, using 1 bit
  if WP is activated then
    if Competition between modes is used then
      Signal the WP mode for the current frame
      1 bit if two competing modes, 2 bits if four competing modes
    end if
    if WP == GL then
      Write  $W$  using 13 bits and  $O$  using 14 bits
    else if WP == GNL then
      Write the 19 ordinates values using the VLC
    else
      for all CTU do
        Signal if the WP is used for the current CTU, using 1 bit
        if WP CTU turn on then
          if WP == LL then
            Write  $W$  using 13 bits and  $O$  using 14 bits
          else if WP == LNL then
            Write the 4 ordinates values using the VLC
          end if
        end if
      end for
    end if
  end if
end for

```

B. CTU Level Activation

For the frame level activation, only the local WPs were decided at a CTU level. For the CTU level activation, this strategy is applied to all WP modes. If a frame meets the constraint (16) each CTU is compressed twice, one time without WP, and a second time with the current WP (GL, GNL, LL, or LNL). The CTU with the best RD performances is then encoded along with its metadata. As for section IV-A,

we put the different WP modes in competition using different combinations. Each competing modes is then used to compress the CTU and the best one in terms of RD performance is selected.

Algorithm 2 CTU Level Activation - HEVC Signaling at the encoder side

```

for all Frame do
  Signal the result of (16) using 1 bit
  if (16) then
    if WP == GL then
      Write the GL WP,  $W$  using 13 bits and  $O$  using 14 bits
    end if
    if WP == GNL then
      Write the GNL WP, 19 ordinates values using the VLC
    end if
    for all CTU do
      if Competition between modes is used then
        Signal the WP mode for the current CTU
        if 1 mode, 1 bit to signal if the WP is used
        if 2 competing modes, 1 bit for no WP or 2 bits for other modes ( 0, 10, 11)
        if 3 competing modes, 2 bits by mode, including no WP (00, 01, 10, 11)
        if 4 competing modes, 1 bit for no WP or 3 bits for other ones (0, 100, 101, 110, 111)
      end if
      if WP local turn on for the CTU then
        if WP == LL then
          Write  $W$  using 13 bits and  $O$  using 14 bits
        else if WP == LNL then
          Write the 19 ordinates values using the VLC
        end if
      end if
    end for
  end if
end for

```

All WPs use the same representation as the one presented in section IV-A, however the signaling syntax changes in some cases, as visible in Algorithm 2. When evaluating only one WP mode, the process is almost similar. For the global WPs, we only add one flag by CTU to signal the activation of the WP, and for the local WPs only the frame decision changes. Indeed, in section IV-A the decision was made after encoding the frame twice, in this case we only evaluate (16). The biggest difference concerns the competition of WP modes, indeed the chosen mode needs to be signaled for each CTU against only once with frame level competition. Since the signaling cost is larger, we evaluated several combinations of WP modes using two, three or the four of them. The detailed signaling for each case is explained in Algorithm 2.

Using 4 WP modes, two levels of activation and different modes of WP competition yield many test cases. The RD performances of all these different strategies are presented in the following section.

V. EXPERIMENTAL RESULTS

To compare the proposed WP methods with the HEVC WP, we used different test sets. The first test set consists of 6 1920x1080 HDR videos of about 200 frames each, taken from the MPEG test set [19], [20], and tone-mapped using 6 different TMOs from the literature, therefore providing 36 SDR sequences. Since this MPEG test set does not include sequences with significant brightness variations interesting for specifically assessing WP methods, we have added a second set of 12 SDR video sequences of about 20 frames with different types of brightness variations (see figure 5) and different resolutions (from SD to HD). These contents were selected because they include quite common brightness variations like fades, concert spot lights or TV shows and could benefit from a better WP. We point out that the selected HDR test set is less representative of the possible illumination changes one may encounter in TV broadcasting for instance. The motivation behind this test set selection is to see if some TMOs create brightness variations at the SDR level that can be compensated by one of the proposed WP modes thus providing better RD performances for these TMOs.

Only the luminance compression is considered in the following results. It could be easily extended to the chrominance components by computing the WP in a similar manner. The WPs were implemented in the HM16.2 using the default lowdelay_P_Main configuration file [21]. For sake of simplicity, we reduce the number of reference frames from 4 to 1. Thus, each P frame only use one reference frame. The same encoding configuration is used in Section V-A, V-B and V-C. The encoding were performed using four quantization parameters (QP) (i.e. QP=22, 27, 32, and 37) and therefore the Bjontegaard rate (BD-rate) gains [22] were computed on the 4 resulting RD points. All the following BD-rate gains are computed using the HM16.2 without WP as a reference. The RD performances of the HEVC reference software WP are reported in each table for comparison purposes.

A. Rate-distortion performances using frame level activation

Table I shows the results using the frame level activation presented in section IV. The first column represents the BD rate gains of the HEVC reference software WP (i.e. with a global linear function) with respect to the HEVC without WP. Here BD-rate gains correspond to negative values. The second one represents the BD-rate gains of the GL WP. In these two cases, the WP uses the same model, therefore, the gains are due to the proposed optical flow based algorithm for WP parameter estimation. For almost each sequence the proposed algorithm outperforms the existing one and on average it provides around 1.6% BD-rate gains. It can be noticed that the current HEVC WP is sometimes responsible of significant loss, especially for the sequences KoreanPop, PopIdol and Guetta0. Actually, these three sequences present mainly local variations which may explain why the HEVC WP fails. The first new WP model, GNL, further improves the RD performances, around 2.5% BD-rate gains. However, for the three same sequences with local variations, we observe some losses compared to the GL algorithm. Indeed, predicting a local transformation with

a global one can be tricky and it's more difficult to find the optimal solution. Figure 4 perfectly illustrates this problem. For the KoreanPop sequence, we can see a large point cloud which is difficult to predict using a global WP. In the local case, the point clouds are narrowed and the local WPs better fit them. These observations are confirmed by the RD results using local WPs for these three sequences. On the other hand, the average RD results are much worse than the ones using global WPs. For the other sequences with global brightness variations, the local WPs are not efficient since it requires a non negligible overhead cost for each CTU.

This first result analysis shows that the global and local WPs are complementary. Using all the WPs in the same encoder could only provide better RD performances, as visible in Table I. The combination of the 4 WP modes provides around 3.13% BD-rate gains compared to the HEVC WP. We tested different combinations using one global and one local WP and the most efficient match is GNL-LL which provides around 2.97% BD-rate gains compared to the HEVC WP. Limiting the frame competition to 2 modes isn't more efficient since the overhead cost savings is negligible. Therefore, for the second test set, we only tested the combination of the 4 WP modes.

To give an insight into the extra complexity, the last line of Table I reports the ratio of the encoding complexity, in percentage, between several WP modes and the reference software. The ratio is averaged over all sequences. It can be observed that the classical HEVC WP allows to reduce the encoding complexity. This can be explained by the fact that with a better quality reference frame, prediction and transform can be avoided (with the skip mode). As for the proposed methods, the complexity increase ranges from 23% to 36%, when only one mode is implemented (GL and GNL). Using the multiple WP modes in competition requires a multi-pass encoding, hence the coding time is multiplied by the number of used WP modes. Note that in all remaining tests of the table, the mode without WP is added to the competition. Therefore, the test labelled LL uses two modes for competition, the test LNL also two modes, 4WP uses five modes, and GL-GNL three modes.

Table II shows the results using the second test set, the 36 SDR tone-mapped sequences. The TMOs mentioned in Table II are the following ones: Mai [3], Reinhard [23], Ferwerda [24], Mertens [25], Raman [26] and Schlick [27]. On average, the observations are the same than for Table I. The non-linear WPs improve the RD performances and the combination of the 4 WP modes provides the best RD performances. Some HDR sequences, like Balloon or Market, present very few brightness variations and therefore presents no improvement in RD performances using the TMOs [3], [23]–[25]. However, for the TMOs [26], [27], these sequences present high RD gains. For the other sequences also, these TMOs usually present the higher RD gains. Indeed, these TMOs can generate temporal flickering on the resulting SDR sequences therefore increasing the SDR temporal variations. The proposed WP modes can predict efficiently these flickering artifacts, therefore compensate it and improve the RD performances.

To conclude on the first results, we can notice that the proposed optical flow based algorithm is better for the estimation

TABLE I
BD-RATE GAINS OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED WP USING FRAME ACTIVATION WITH RESPECT TO THE HEVC STANDARD WITHOUT ANY WP. SDR TEST SEQUENCES IN A LOW DELAY P CONFIGURATION USING 1 FRAME OF TYPE P

Sequences	HEVC WP	GL	GNL	LL	LNL	4 WP	GNL LL
BodyPainting	-9,15%	-10,26%	-11,13%	-10,65%	-9,99%	-12,76%	-13,22%
Concert	-2,35%	-2,38%	-2,44%	-0,59%	-1,48%	-2,96%	-2,49%
Eurovision	0,01%	0,01%	0,01%	0,01%	-0,15%	0,01%	0,01%
KoreanPop	8,17%	-3,02%	-0,79%	-4,49%	-3,30%	-4,62%	-4,51%
LeMatch	0,04%	0,04%	0,04%	0,12%	-0,06%	-0,15%	0,12%
Guetta-0	3,65%	-0,30%	-0,27%	-0,26%	-0,06%	-0,52%	-0,44%
Guetta-2	-0,37%	-0,06%	-0,10%	-0,16%	0,02%	-0,18%	-0,29%
RedBull	-37,08%	-36,20%	-40,91%	-4,03%	-6,86%	-40,65%	-40,70%
PopIdol	0,61%	-1,34%	-0,54%	-2,07%	-1,35%	-2,07%	-2,23%
SkyFade	-12,60%	-13,63%	-13,66%	-0,06%	-2,26%	-13,81%	-12,88%
SunFade	-23,00%	-24,17%	-25,72%	5,89%	4,63%	-24,73%	-24,59%
Tangerine	-10,27%	-10,30%	-16,64%	3,51%	0,74%	-17,55%	-16,72%
Average BD-Rate Gain	-6,86%	-8,47%	-9,34%	-1,07%	-1,68%	-9,99%	-9,83%
Average encoding complexity	97,11 %	123,42 %	136,39 %	233,98 %	-225,48 %	588,26 %	326,43 %

of WP model parameters than the actual one in the HM16.2, and that the proposed non-linear WPs outperforms the linear ones. We also demonstrated that global and local weighted prediction are complementary and their combination further improves the RD performances.

B. Rate-distortion performances using CTU level activation

Table III shows the results using the CTU level activation presented in section IV. On average, the observations for the frame level activation remains valid for this case. The principal difference with the previous results of Table I is that each mode needs an overhead cost for each CTU. Surprisingly, it does not impact the RD performances of global WPs. On the contrary, the GL and GNL WPs using CTU activation present 0.43% BD-rate gains and 0.37% BD-rate gains compared to the ones using frame activation. Note that, for GL and GNL, the overhead cost is relatively small (1 bit per CTU). As explained in section IV, the process for local WPs using CTU activation is almost similar to the ones using frame activation. In the CTU activation case, we rely on the threshold (16) to decide the use of the WP.

Using the combination of 4 WP at CTU level provides the same results as those obtained at frame level. The better accuracy of this competition just counterbalance the overhead cost for each CTU. Also here, we tested several combinations of WP, as explained in section IV. Table III only presents the RD performance of the most efficient combination, which is GL-GNL-LL WPs, providing on average 3.44% BD-rate gains against the classical HEVC WP. Indeed, limiting the competition to these three modes is more efficient than using the 4 WP model since we move from a cost of 3 bits by CTU to 2 bits by CTU (See algorithm 2).

Table IV shows the results using the same test set as in Table II. On average, observations are coherent with the ones made in Table II or III, except that the competition of four WP modes provides the best RD performances instead of GL-GNL-LL. The proposed WP modes outperforms the RD performances of

the existing WP modes and the better RD gains are achieved with the TMOs that generate temporal flickering artifacts.

To conclude on these results, we showed that using a CTU competition for WPs modes can be more interesting than the frame level activation except for some marginal cases where the overhead cost limits the BD-rate gains.

C. HDR rate-distortion performances using best WP modes

The previous sections V-A and V-B present the RD gains on SDR sequences. In this section, we present the RD performances obtained on the HDR level using the invertible TMO [3] and the best WP modes: the 4 WP modes with frame level activation, the 4WP modes and GL-GNL-LL modes with CTU level activation. The SDR RD performances of this configuration are already presented in Table II and IV. In this case, the SDR videos are inverse tone mapped to reconstruct the HDR sequences and the RD performances are computed using different HDR metrics: the tPSNR [28] in Table V, the PU-PSNR [29] in Table VI, the PU-SSIM [29] in Table VII and the HDRVDP [30] in Table VIII.

On average, the RD gains are consistent with all metrics, a little lower for the HDRVDP. These RD performances are consistent with the TMO-generated SDR ones and demonstrate that improving the SDR RD performances can improve the HDR ones. Furthermore, the SDR content is not impacted and therefore the temporal quality is preserved.

D. Rate-distortion performances using multiple references

The previous results were presented using the default HEVC lowdelay_P_Main configuration file with only one reference frame. This configuration allows to more accurately assess the contribution of each WP models. Indeed, with multiple references frames, local blocks can be predicted using different reference frames, thus allowing implicit local brightness adaptation when using only global WP. For multiple references, we have chosen the low delay B configuration instead of the low delay P configuration. Although the two configurations

TABLE II

BD-RATE GAINS OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED WP USING FRAME ACTIVATION WITH RESPECT TO THE HEVC STANDARD WITHOUT ANY WP. SDR TONE-MAPPED TEST SEQUENCES IN A LOW DELAY P CONFIGURATION USING 1 FRAME OF TYPE P

Sequences	TMO	HEVC WP	GL	GNL	LL	LNL	4 WP
Balloon	Mai	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
	Reinhard	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
	Ferwerda	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
	Mertens	0.01 %	0.02 %	0.24 %	-0.04 %	-0.15 %	-0.12 %
	Raman	-5.95 %	-5.74 %	-7.88 %	-2.30 %	-3.97 %	-7.52 %
	Schlick	0.05 %	0.05 %	0.05 %	0.05 %	0.05 %	0.05 %
Carousel3	Mai	-0.03 %	-0.13 %	-0.01 %	0.25 %	-0.17 %	-0.13 %
	Reinhard	-0.32 %	-0.40 %	-0.09 %	-0.43 %	-0.49 %	-0.61 %
	Ferwerda	-0.07 %	-0.31 %	-0.03 %	-0.15 %	-0.40 %	-0.28 %
	Mertens	-0.07 %	0.01 %	-0.01 %	-0.33 %	-0.23 %	-0.28 %
	Raman	-4.00 %	-3.58 %	-3.53 %	-2.11 %	-2.15 %	-3.52 %
	Schlick	-0.07 %	0.02 %	0.01 %	-0.37 %	-0.34 %	-0.43 %
Carousel4	Mai	0.00 %	-0.22 %	-0.14 %	-0.33 %	-0.53 %	-0.60 %
	Reinhard	-0.04 %	-0.47 %	-0.37 %	-0.53 %	-0.80 %	-0.86 %
	Ferwerda	-0.22 %	-0.79 %	-0.47 %	-1.06 %	-1.12 %	-1.28 %
	Mertens	0.00 %	-0.02 %	0.04 %	-0.46 %	-0.42 %	-0.56 %
	Raman	-3.00 %	-2.94 %	-3.26 %	-1.80 %	-2.09 %	-3.06 %
	Schlick	-0.06 %	-0.36 %	-0.28 %	-0.65 %	-0.99 %	-1.00 %
FireEater	Mai	-0.10 %	-0.12 %	-0.10 %	0.06 %	0.01 %	-0.37 %
	Reinhard	0.06 %	0.01 %	0.14 %	-0.19 %	-0.24 %	-0.18 %
	Ferwerda	-0.67 %	-0.69 %	-0.97 %	-1.97 %	-0.64 %	-2.16 %
	Mertens	-0.01 %	0.00 %	-0.01 %	-0.24 %	-0.11 %	-0.25 %
	Raman	-5.80 %	-5.34 %	-4.92 %	-2.82 %	-2.63 %	-5.28 %
	Schlick	-2.42 %	-2.53 %	-5.25 %	-3.15 %	-2.86 %	-5.41 %
Market	Mai	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
	Reinhard	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
	Ferwerda	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
	Mertens	-0.05 %	-0.04 %	0.05 %	-0.27 %	-0.45 %	-0.42 %
	Raman	-30.22 %	-25.90 %	-32.84 %	-17.25 %	-21.83 %	-31.51 %
	Schlick	-4.84 %	-13.79 %	-31.02 %	-16.01 %	-27.68 %	-37.33 %
Tibul	Mai	-0.01 %	-1.62 %	-1.29 %	-1.58 %	-3.13 %	-3.33 %
	Reinhard	-0.15 %	-2.03 %	-2.53 %	-2.92 %	-4.50 %	-4.71 %
	Ferwerda	-0.38 %	-1.45 %	-2.67 %	-4.75 %	-5.20 %	-5.76 %
	Mertens	-0.25 %	-0.46 %	-0.31 %	-2.78 %	-2.88 %	-3.06 %
	Raman	-6.12 %	-6.12 %	-9.77 %	-6.03 %	-7.36 %	-9.17 %
	Schlick	-1.51 %	-2.28 %	-5.28 %	-5.67 %	-7.59 %	-7.60 %
Average	Mai	-0.02 %	-0.35 %	-0.25 %	-0.26 %	-0.63 %	-0.74 %
	Reinhard	-0.07 %	-0.48 %	-0.47 %	-0.68 %	-1.00 %	-1.06 %
	Ferwerda	-0.22 %	-0.54 %	-0.69 %	-1.32 %	-1.22 %	-1.58 %
	Mertens	-0.06 %	-0.08 %	0.00 %	-0.69 %	-0.71 %	-0.78 %
	Raman	-9.18 %	-8.27 %	-10.37 %	-5.39 %	-6.67 %	-10.01 %
	Schlick	-1.48 %	-3.15 %	-6.96 %	-4.30 %	-6.57 %	-8.62 %
Average		-1.85 %	-2.14 %	-3.12 %	-2.11 %	-2.80 %	-3.80 %

are similar, the low delay B configuration enables the use of bi-prediction for all frames. Hence, this configuration allows to evaluate the impact of both, multiple-referencing and bi-prediction, on the proposed scheme efficiency. The weighted bi-prediction add another candidate for each local block, again implicitly accounting for local illumination change compensation.

To assess the performance of the proposed optical flow based algorithm and new WP models, we also performed tests using the default HEVC lowdelay_B_Main configuration file while maintaining the 4 reference frames. Table IX shows the results using the frame level activation and Table X, the results with CTU level activation. For these tests, we only focus on the SDR test set.

With the GL results in Table IX, we can see that the proposed optical flow based algorithm still outperforms the HEVC algorithm with a BD-rate improvement of 3%. However, unlike the configuration using one reference frame, the non linear WP models do not outperform the linear one. This observation is mainly explained with the overhead difference between the linear WP model and the non linear ones. Indeed, the non linear WP overhead is multiplied by four since we use four reference frames. We also tested several competitions of WP model and using the GL and GNL models, we obtain a BD-rate improvement of 0.5% when compared to the GL model.

For the CTU level results in Table X, we can notice that the GNL WP model outperform the GL one with a BD-rate gain around 0.5%. For the local WP models, the LL WP model

TABLE III

BD-RATE GAINS OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED WP USING CTU ACTIVATION WITH RESPECT TO THE HEVC STANDARD WITHOUT ANY WP. SDR TEST SEQUENCES IN A LOW DELAY P CONFIGURATION USING 1 FRAME OF TYPE P

Sequences	HEVC WP	GL	GNL	LL	LNL	4 WP	GL GNL LL
BodyPainting	-9,15%	-12,12%	-12,58%	-10,60%	-9,88%	-14,46%	-14,45%
Concert	-2,35%	-4,48%	-4,38%	-0,60%	-1,52%	-4,88%	-5,17%
Eurovision	0,01%	0,01%	0,01%	0,01%	0,01%	0,01%	0,01%
KoreanPop	8,17%	-5,54%	-4,54%	-4,49%	-3,30%	-7,35%	-7,28%
LeMatch	0,04%	-0,14%	-0,15%	0,12%	-0,16%	-0,24%	-0,27%
Guetta-0	3,65%	-0,64%	-0,61%	-0,26%	-0,08%	-1,22%	-1,12%
Guetta-2	-0,37%	-0,42%	-0,60%	-0,16%	-0,04%	-0,95%	-0,78%
RedBull	-37,08%	-35,05%	-39,02%	-3,92%	-6,24%	-36,49%	-38,68%
PopIdol	0,61%	-2,44%	-2,24%	-2,07%	-1,36%	-3,99%	-3,85%
SkyFade	-12,60%	-12,98%	-13,54%	0,16%	-1,83%	-14,21%	-15,25%
SunFade	-23,00%	-21,43%	-22,95%	6,13%	4,51%	-21,31%	-21,80%
Tangerine	-10,27%	-11,55%	-15,93%	4,01%	2,03%	-14,50%	-14,95%
Average	-6,86%	-8,90%	-9,71%	-0,97%	-1,49%	-9,97%	-10,30%

provides better results than the LNL WP model. As in Section V-B, the 4 WP competition is penalized by a signaling cost of 3 bits by CTU while the competition GL, GNL and LL only uses 2 bits by CTU. For these reasons, only the competition of GL, GNL and LL provides BD-rate gains, around 0,8%, when compared to the GL WP model.

To assess the performance of the proposed optical flow based WP algorithm, we also performed tests using the default HEVC Random Access configuration file with 8 reference frames. Table XI shows the results using the frame level activation and Table XII, the results with CTU level activation. The Random Access configuration provides lower BD-rate gains than the Low Delay configuration (compare Table IX with Table XI, and X with XII).

Interestingly, the BD-rate gains of all WP methods, including the HEVC one, are reduced compared to low delay configurations (compare Table IX with Table XI, and X with XII). This can be explained by the fact that in the Random Access configuration, future frames are used as reference. Now, if an illumination change occurs on the current frame, it is highly possible that the future reference frames experience the same illumination. Consequently WP, which aims at compensating for illumination changes, will bring little gain in the Random Access configuration. However, all the proposed WP methods provide larger gains than the HEVC WP.

E. Localized rate-distortion performances

Evaluating the RD performances of the WP in a fair manner is not obvious. For some sequences, the gains can be localized on a specific frame and therefore minimized when averaging the gains over the entire sequence. To overcome this problem, we computed the BD-rate gains for each frame, using 4 RD points (4 QPs), for each frame and plotted them on Figure 6. The green curves represent the best competition mode using frame competition and the blue curves represent the best ones using the CTU competition. The abscissa represents the frame index in display order of the sequence and the ordinates, the associated BD-Rate gains. For the BodyPainting sequence, limiting the test sequence to the frame 4 to 9, the BD-rate gains would have been even better.

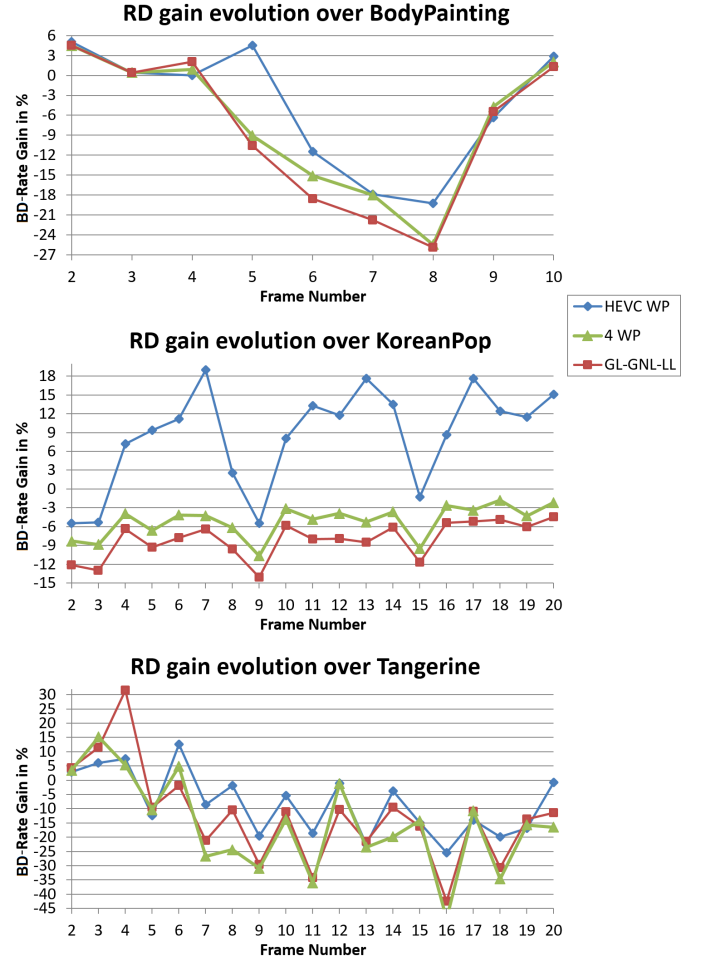


Fig. 6. Rate Gain by frame using SDR PSNR and SDR rate. From top to bottom: BodyPainting, Tangerine, KoreanPop

TABLE IV

BD-RATE GAINS OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED WP USING CTU ACTIVATION WITH RESPECT TO THE HEVC STANDARD WITHOUT ANY WP. SDR TONE-MAPPED TEST SEQUENCES IN A LOW DELAY P CONFIGURATION USING 1 FRAME OF TYPE P

Sequences	TMO	HEVC WP	GL	GNL	LL	LNL	4 WP	GL GNL LL
Balloon	Mai	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
	Reinhard	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
	Ferwerda	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
	Mertens	0.01 %	-0.09 %	-0.14 %	-0.04 %	-0.12 %	-0.33 %	-0.24 %
	Raman	-5.95 %	-5.72 %	-7.59 %	-2.30 %	-4.01 %	-7.74 %	-7.73 %
	Schlick	0.05 %	0.05 %	0.05 %	0.05 %	0.05 %	0.05 %	0.05 %
Carousel3	Mai	-0.03 %	-0.58 %	-0.59 %	0.15 %	-0.14 %	-1.04 %	-0.84 %
	Reinhard	-0.32 %	-0.82 %	-0.88 %	-0.43 %	-0.51 %	-1.67 %	-1.58 %
	Ferwerda	-0.07 %	-1.17 %	-0.99 %	-0.16 %	-0.36 %	-1.91 %	-1.55 %
	Mertens	-0.07 %	-0.37 %	-0.32 %	-0.33 %	-0.21 %	-0.77 %	-0.73 %
	Raman	-4.00 %	-4.06 %	-4.16 %	-2.08 %	-2.04 %	-5.29 %	-5.20 %
	Schlick	-0.07 %	-0.60 %	-0.52 %	-0.37 %	-0.36 %	-1.04 %	-1.03 %
Carousel4	Mai	0.00 %	-0.83 %	-0.76 %	-0.33 %	-0.53 %	-1.57 %	-1.38 %
	Reinhard	-0.04 %	-1.16 %	-1.20 %	-0.52 %	-0.79 %	-2.07 %	-1.88 %
	Ferwerda	-0.22 %	-1.63 %	-1.52 %	-1.06 %	-1.12 %	-2.80 %	-2.62 %
	Mertens	0.00 %	-0.60 %	-0.57 %	-0.45 %	-0.42 %	-1.36 %	-1.18 %
	Raman	-3.00 %	-3.59 %	-4.01 %	-1.79 %	-2.07 %	-4.90 %	-4.76 %
	Schlick	-0.06 %	-1.13 %	-1.05 %	-0.65 %	-0.95 %	-2.18 %	-1.89 %
FireEater	Mai	-0.10 %	-0.87 %	-0.62 %	0.06 %	0.01 %	-1.62 %	-1.22 %
	Reinhard	0.06 %	-0.14 %	-0.23 %	-0.19 %	-0.04 %	-1.25 %	-0.81 %
	Ferwerda	-0.67 %	-2.04 %	-2.03 %	-1.97 %	-1.51 %	-4.75 %	-3.85 %
	Mertens	-0.01 %	-0.28 %	-0.23 %	-0.24 %	-0.13 %	-0.66 %	-0.59 %
	Raman	-5.80 %	-5.90 %	-5.70 %	-2.82 %	-2.62 %	-7.52 %	-7.32 %
	Schlick	-2.42 %	-2.78 %	-5.71 %	-3.15 %	-2.86 %	-6.70 %	-6.39 %
Market	Mai	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
	Reinhard	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
	Ferwerda	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %	0.01 %
	Mertens	-0.05 %	-0.34 %	-0.35 %	-0.27 %	-0.38 %	-0.95 %	-0.70 %
	Raman	-30.22 %	-26.35 %	-31.27 %	-17.26 %	-21.32 %	-31.60 %	-31.88 %
	Schlick	-4.84 %	-16.10 %	-38.67 %	-16.00 %	-27.92 %	-39.73 %	-39.48 %
Tibul	Mai	-0.01 %	-3.93 %	-3.19 %	-1.58 %	-3.11 %	-5.82 %	-5.29 %
	Reinhard	-0.15 %	-3.90 %	-4.31 %	-2.92 %	-4.46 %	-7.13 %	-6.49 %
	Ferwerda	-0.38 %	-3.27 %	-4.78 %	-4.75 %	-5.20 %	-8.73 %	-8.13 %
	Mertens	-0.25 %	-1.71 %	-1.69 %	-2.78 %	-2.88 %	-4.60 %	-4.03 %
	Raman	-6.12 %	-7.40 %	-11.04 %	-6.05 %	-7.04 %	-12.21 %	-12.09 %
	Schlick	-1.51 %	-3.76 %	-7.14 %	-5.67 %	-7.54 %	-9.94 %	-8.99 %
Average	Mai	-0.02 %	-1.03 %	-0.86 %	-0.28 %	-0.63 %	-1.67 %	-1.45 %
	Reinhard	-0.07 %	-1.00 %	-1.10 %	-0.67 %	-0.96 %	-2.02 %	-1.79 %
	Ferwerda	-0.22 %	-1.35 %	-1.55 %	-1.32 %	-1.36 %	-3.03 %	-2.69 %
	Mertens	-0.06 %	-0.57 %	-0.55 %	-0.69 %	-0.69 %	-1.45 %	-1.25 %
	Raman	-9.18 %	-8.84 %	-10.63 %	-5.38 %	-6.52 %	-11.54 %	-11.50 %
	Schlick	-1.48 %	-4.05 %	-8.84 %	-4.30 %	-6.60 %	-9.92 %	-9.62 %
Average		-1.85 %	-2.81 %	-3.92 %	-2.11 %	-2.79 %	-4.94 %	-4.72 %

TABLE V

HDR BD-RATE GAINS USING TPSNR OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED WP WITH RESPECT TO THE HEVC STANDARD WITHOUT ANY WP. HDR TEST SEQUENCES IN A LOW DELAY P CONFIGURATION USING 1 FRAME OF TYPE P

Sequences	TMO	HEVC WP	Frame 4WP	CTU 4WP	CTU GL GNL LL
Balloon	Mai	0.01 %	0.01 %	0.01 %	0.01 %
Carousel3	Mai	-0.10 %	-0.14 %	-0.91 %	0.39 %
Carousel4	Mai	0.05 %	-0.63 %	-1.56 %	-1.34 %
FireEater	Mai	-0.15 %	-0.33 %	-1.94 %	-1.57 %
Market	Mai	0.01 %	0.01 %	0.01 %	0.01 %
Tibul	Mai	-0.05 %	-3.24 %	-5.83 %	-5.18 %
Average		-0.04 %	-0.72 %	-1.70 %	-1.41 %

TABLE VI

HDR BD-RATE GAINS USING PU-PSNR OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED WP WITH RESPECT TO THE HEVC STANDARD WITHOUT ANY WP. HDR TEST SEQUENCES IN A LOW DELAY P CONFIGURATION USING 1 FRAME OF TYPE P

Sequences	TMO	HEVC WP	Frame 4WP	CTU 4WP	CTU GL GNL LL
Balloon	Mai	0.01 %	0.01 %	0.01 %	0.01 %
Carousel3	Mai	-0.08 %	-0.02 %	-0.87 %	-0.34 %
Carousel4	Mai	0.07 %	-0.62 %	-1.53 %	-1.31 %
FireEater	Mai	-0.12 %	-0.34 %	-1.95 %	-1.63 %
Market	Mai	0.01 %	0.01 %	0.01 %	0.01 %
Tibul	Mai	-0.03 %	-3.25 %	-5.80 %	-5.15 %
Average		-0.02 %	-0.70 %	-1.69 %	-1.40 %

TABLE VII

HDR BD-RATE GAINS USING PU-SSIM OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED WP WITH RESPECT TO THE HEVC STANDARD WITHOUT ANY WP. HDR TEST SEQUENCES IN A LOW DELAY P CONFIGURATION USING 1 FRAME OF TYPE P

Sequences	TMO	HEVC WP	Frame 4WP	CTU 4WP	CTU GL GNL LL
Balloon	Mai	0.01 %	0.01 %	0.01 %	0.01 %
Carousel3	Mai	0.34 %	0.22 %	-0.72 %	-0.08 %
Carousel4	Mai	0.16 %	-0.39 %	-1.45 %	-1.12 %
FireEater	Mai	-0.13 %	-0.21 %	-1.55 %	-1.36 %
Market	Mai	0.02 %	0.02 %	0.02 %	0.02 %
Tibul	Mai	-0.03 %	-2.76 %	-5.28 %	-4.89 %
Average		0.06 %	-0.52 %	-1.50 %	-1.24 %

TABLE VIII

HDR BD-RATE GAINS USING HDRVDP OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED WP WITH RESPECT TO THE HEVC STANDARD WITHOUT ANY WP. HDR TEST SEQUENCES IN A LOW DELAY P CONFIGURATION USING 1 FRAME OF TYPE P

Sequences	TMO	HEVC WP	Frame 4WP	CTU 4WP	CTU GL GNL LL
Balloon	Mai	0.01 %	0.01 %	0.01 %	0.01 %
Carousel3	Mai	0.17 %	0.45 %	0.46 %	1.12 %
Carousel4	Mai	0.33 %	-0.78 %	-2.04 %	-1.56 %
FireEater	Mai	0.47 %	-0.29 %	-1.29 %	-1.37 %
Market	Mai	0.01 %	0.01 %	0.01 %	0.01 %
Tibul	Mai	-0.02 %	-3.02 %	-5.21 %	-4.55 %
Average		0.16 %	-0.60 %	-1.34 %	-1.06 %

On average, the proposed WPs always outperforms the HEVC weighted prediction except in a specific case at the beginning of the Tangerine sequence using the combination of GL-GNL-LL. For this frame, the threshold (16) wrongly activates the WP for the current frame therefore imposing an overhead cost for all CTUs. Since we are at the beginning of a fade-in, the corresponding rate of the frame is very low therefore the overhead cost result in a large BD-rate loss, while in fact it's negligible compared to the other frame rates. This presentation for the BD-rate gains also allows to appreciate the performance of the WP. Indeed, for some frame, the BD-rate gains go up to 26% for BodyPainting, 14% for KoreanPop and 45% for Tangerine.

Figure 7 shows the BD-rate gains computed on the HDR sequences Tibul, FireEater and Carousel4 using the HDRVDP metric. In this case, we show the results using the 4 WP modes with CTU competition since it provides the best RD performances in Tables II and IV. The HDR sequences contain around 200 frames, but for sake of clarity we only plotted the gains for a subset of frame with much temporal variations. As explain above, encoding all the frame of the sequence can round down the WP gains, thereby Figure 7 shows BD-Rate gains up to 14% for Tibul, 21% for FireEater and 7% for Carousel4.

F. Learning the weights locally to reduce the overhead

Local methods need to send the WP coefficients for each block. To avoid this overhead, the parameters can be computed on neighboring blocks as in [15]–[17]. This, however, increases the encoding cost of the residue due to the lack of

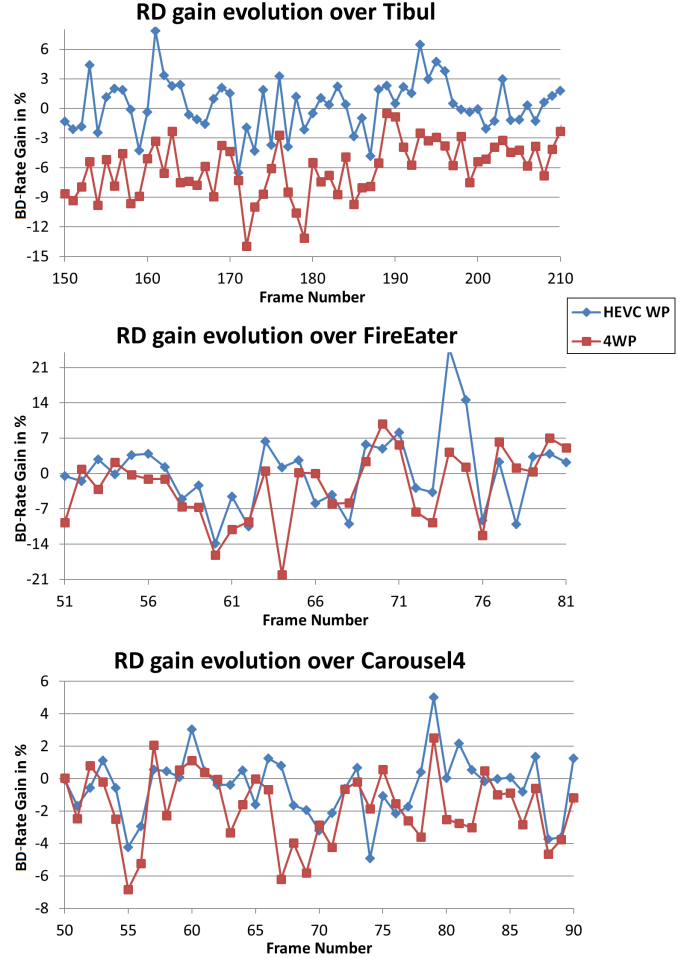


Fig. 7. Rate Gain by frame using HDRVDP and SDR rate. From top to bottom: Tibul, FireEater, Carousel4

accuracy of the WP parameters. Thus, this leads to a tradeoff. This idea is compatible with the proposed weight computation based on optical flow and has been tested for LL (see Table XIII) and LNL (see Table XIV) WP methods. Different context sizes have been tested to compute the coefficients. The contexts are either 1, 4 or 64 pixels large. In all cases, learning the weights from a local and causal neighborhood achieves a greater BD-rate gain than the method with signaling. The best performance is obtained when the context is 64 pixel large for LL WP and 4 pixel large for LNL WP.

G. Discussions

For sake of comparison, we implemented the solution proposed in [5]. Our implementation tested with various parameters and same test conditions did not lead to any RD gains compared to [3], while the proposed WPs improve the BD-Rate performances of the single layer compression scheme [3].

As explained in section IV the optical flows and WPs are computed prior to the encoding process. The accuracy of WPs could possibly be improved if done directly in the encoding loop. Indeed, all computation could be made with the reconstructed frame S_{t-1} instead of the source frame. Besides,

TABLE IX

BD-RATE GAINS OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED WP USING FRAME ACTIVATION WITH RESPECT TO THE HEVC STANDARD WITHOUT ANY WP. SDR TEST SEQUENCES IN A LOW DELAY B CONFIGURATION USING 4 FRAMES OF TYPE B

Sequences	HEVC WP	GL	GNL	LL	LNL	4 WP	GL GNL
BodyPainting	-13,58%	-14,37%	-13,60%	-12,14%	-10,97%	-14,87%	-15,15%
Concert	-18,28%	-18,29%	-17,55%	-5,04%	-3,89%	-18,30%	-18,35%
Eurovision	0,02%	0,01%	-0,13%	0,10%	-0,40%	-0,68%	-0,14%
KoreanPop	2,69%	-4,77%	-3,26%	-3,59%	-2,17%	-3,21%	-4,85%
LeMatch	0,12%	-0,11%	-0,11%	0,10%	0,19%	0,15%	-0,10%
Guetta-0	2,95%	-0,67%	-0,63%	-0,79%	-0,16%	-0,70%	-0,94%
Guetta-2	-0,15%	-0,48%	-0,52%	-0,35%	-0,26%	-0,56%	-0,66%
RedBull	-42,81%	-50,55%	-50,45%	-2,28%	-0,31%	-50,49%	-51,09%
PopIdol	0,46%	-2,36%	-2,21%	-2,08%	-1,27%	-2,24%	-2,92%
SkyFade	-21,30%	-23,37%	-23,13%	-0,34%	1,09%	-23,66%	-24,34%
SunFade	-38,17%	-44,89%	-41,28%	6,98%	8,64%	-42,95%	-44,46%
Tangerine	-17,99%	-21,84%	-22,86%	4,15%	3,60%	-25,15%	-24,81%
Average	-12,17%	-15,14%	-14,64%	-1,27%	-0,49%	-15,22%	-15,66%

TABLE X

BD-RATE GAINS OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED WP USING CTU ACTIVATION WITH RESPECT TO THE HEVC STANDARD WITHOUT ANY WP. SDR TEST SEQUENCES IN A LOW DELAY B CONFIGURATION USING 4 FRAMES OF TYPE B

Sequences	HEVC WP	GL	GNL	LL	LNL	4 WP	GL GNL LL
BodyPainting	-13,58%	-13,68%	-13,55%	-12,07%	-10,88%	-14,78%	-15,56%
Concert	-18,28%	-17,41%	-17,17%	-4,96%	-3,78%	-16,89%	-18,63%
Eurovision	0,02%	-0,21%	-0,53%	0,29%	-0,34%	-1,19%	-0,54%
KoreanPop	2,69%	-6,47%	-6,37%	-3,59%	-2,17%	-7,13%	-8,37%
LeMatch	0,12%	0,54%	0,81%	0,53%	0,25%	0,27%	1,15%
Guetta-0	2,95%	-0,84%	-1,06%	-0,61%	-0,02%	-1,50%	-1,86%
Guetta-2	-0,15%	-0,67%	-1,20%	-0,24%	-0,05%	-1,13%	-1,33%
RedBull	-42,81%	-45,22%	-47,14%	-0,17%	-0,20%	-42,56%	-46,83%
PopIdol	0,46%	-3,03%	-3,50%	-2,07%	-1,27%	-4,25%	-4,94%
SkyFade	-21,30%	-19,98%	-21,21%	0,99%	1,10%	-17,43%	-21,38%
SunFade	-38,17%	-34,94%	-33,75%	9,28%	9,55%	-26,91%	-31,08%
Tangerine	-17,99%	-17,62%	-19,81%	5,60%	5,19%	-17,19%	-19,13%
Average	-12,17%	-13,29%	-13,71%	-0,59%	-0,22%	-12,56%	-14,04%

TABLE XI

BD-RATE GAINS OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED WP USING FRAME ACTIVATION WITH RESPECT TO THE HEVC STANDARD WITHOUT ANY WP. SDR TEST SEQUENCES IN A RANDOM ACCESS CONFIGURATION USING 8 FRAMES OF TYPE B

Sequences	HEVC WP	GL	GNL	LL	LNL	4 WP	GL GNL
BodyPainting	-18,02%	-18,93 %	-17,95 %	-15,38 %	-13,63 %	-19,20 %	-19,68 %
Concert	-22,68%	-22,54 %	-21,67 %	-5,52 %	-3,90 %	-22,17 %	-22,55 %
Eurovision	0,20 %	0,01 %	-0,10 %	-0,21 %	-0,18 %	-6,27 %	-1,19 %
KoreanPop	4,31 %	-3,50 %	-3,13 %	-3,02 %	-2,55 %	-3,61 %	-3,55 %
LeMatch	0,07 %	0,08 %	0,03 %	0,79 %	1,04 %	0,12 %	0,16 %
Guetta-0	2,75 %	-0,48 %	-0,49 %	-0,01 %	0,78 %	-0,12 %	-0,66 %
Guetta-2	0,47 %	-0,39 %	-0,42 %	-0,28 %	-0,10 %	-0,58 %	-0,71 %
RedBull	-32,91%	-38,71 %	-38,68 %	-1,02 %	-0,69 %	-38,28 %	-39,05 %
PopIdol	-0,56 %	-3,03 %	-2,80 %	-3,27 %	-2,48 %	-3,27 %	-3,80 %
SkyFade	-11,40%	-12,36 %	-12,28 %	0,55 %	1,51 %	-12,13 %	-12,81 %
SunFade	-24,67%	-28,87 %	-26,58 %	5,24 %	6,51 %	-27,23 %	-28,51 %
Tangerine	-12,37%	-14,87 %	-15,62 %	3,59 %	3,41 %	-16,76 %	-16,84 %
Average	-9,57%	-11,97%	-11,64%	-1,55%	-0,86%	-12,46%	-12,43%

regarding the optical flow, we tested a few others [31], [32] but the preliminary results were better using the presented one. However, it's quite likely that more robust optical flows could improve the precision of the weighted prediction and therefore the RD performances.

VI. CONCLUSION

In the context of the backward compatible HDR compression of videos, we demonstrated that using a weighted prediction is better than using a temporally optimized TMO in terms of RD performances. This is explained by the fact that, with the weighted prediction, the joint optimization problem of

TABLE XII

BD-RATE GAINS OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED WP USING CTU ACTIVATION WITH RESPECT TO THE HEVC STANDARD WITHOUT ANY WP. SDR TEST SEQUENCES IN A RANDOM ACCESS CONFIGURATION USING 8 FRAMES OF TYPE B

Sequences	HEVC WP	GL	GNL	LL	LNL	4 WP	GL GNL LL
BodyPainting	-18,02 %	-18,12%	-17,97 %	-15,60 %	-13,44 %	-19,24 %	-19,83%
Concert	-22,68 %	-21,46%	-21,28 %	-5,73 %	-3,69 %	-20,58 %	-22,29%
Eurovision	0,20 %	-0,10%	-0,24 %	0,21 %	0,01 %	-0,87 %	-0,64%
KoreanPop	4,31 %	-5,23%	-5,29 %	-2,33 %	-1,48 %	-6,05 %	-7,58%
LeMatch	0,07 %	0,44%	0,47 %	0,72 %	0,29 %	0,53 %	0,48%
Guetta-0	2,75 %	-0,65%	-0,97 %	-0,15 %	0,48 %	-1,02 %	-0,91%
Guetta-2	0,47 %	-0,37%	-0,56 %	-0,11 %	0,13 %	-0,73 %	-0,85%
RedBull	-32,91 %	-34,62%	-35,22 %	-0,19 %	-0,35 %	-33,34 %	-35,17%
PopIdol	-0,56 %	-2,84%	-3,29 %	-1,95 %	-1,55 %	-4,57 %	-4,86%
SkyFade	-11,40 %	-10,56%	-11,34 %	0,95 %	1,59 %	-8,95 %	-10,62%
SunFade	-24,67 %	-22,45%	-21,80 %	7,42 %	8,17 %	-17,02 %	-19,26%
Tangerine	-12,37 %	-12,98%	-13,61 %	4,27 %	4,57 %	-12,45 %	-13,33%
Average	-9,57 %	-10,75%	-10,92%	-1,04%	-0,44%	-10,36%	-11,24%

TABLE XIII

BD-RATE GAINS OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED LL WP WITHOUT SIGNALLING USING A DIFFERENT NEIGHBORHOOD WIDTH (NBR). SDR TEST SEQUENCES IN A LOW DELAY P CONFIGURATION USING 1 FRAME OF TYPE P

Sequences	HEVC WP	LL as in Tab I	Without signalling NBR 1	Without signalling NBR 4	Without signalling NBR 64
BodyPainting	-9,15%	-10,65 %	-11,65 %	-11,76 %	-12,05 %
Concert	-2,35%	-0,59 %	-2,88 %	-3,03 %	-2,86 %
Eurovision	0,01%	0,01 %	0,01 %	0,01 %	0,01 %
KoreanPop	8,17%	-4,49 %	-5,11 %	-5,21 %	-5,25 %
LeMatch	0,04%	0,12 %	0,08 %	0,18 %	0,04 %
Guetta-0	3,65%	-0,33 %	-0,83 %	-0,81 %	-0,70 %
Guetta-2	-0,37%	-0,16 %	-0,07 %	-0,16 %	-0,10 %
RedBull	-37,08%	-4,03 %	-14,18 %	-14,40 %	-14,84 %
PopIdol	0,61%	-2,07 %	-2,55 %	-2,79 %	-2,65 %
SkyFade	-12,60%	-0,06 %	-3,23 %	-4,33 %	-4,62 %
SunFade	-23,00%	5,89 %	0,90 %	0,34 %	0,02 %
Tangerine	-10,27%	3,51 %	-0,94 %	-1,63 %	-1,97 %
Average	-6,86%	-1,07%	-3,37%	-3,63%	-3,75%

TABLE XIV

BD-RATE GAINS OF THE HEVC WP (COLUMN 1) AND OF THE PROPOSED LNL WP WITHOUT SIGNALLING USING A DIFFERENT NEIGHBORHOOD WIDTH (NBR). SDR TEST SEQUENCES IN A LOW DELAY P CONFIGURATION USING 1 FRAME OF TYPE P

Sequences	HEVC WP	LNL as in Tab I	Without signalling NBR 1	Without signalling NBR 4	Without signalling NBR 64
BodyPainting	-9,15%	-9,99 %	-9,81 %	-10,74 %	-12,00 %
Concert	-2,35%	-1,48 %	-3,40 %	-3,73 %	-3,30 %
Eurovision	0,01%	-0,15 %	0,01 %	0,01 %	0,01 %
KoreanPop	8,17%	-3,30 %	-3,77 %	-3,82 %	-3,74 %
LeMatch	0,04%	-0,06 %	-0,47 %	-0,34 %	-0,16 %
Guetta-0	3,65%	-0,06 %	-0,41 %	-0,43 %	-0,37 %
Guetta-2	-0,37%	0,02 %	-0,11 %	-0,02 %	-0,09 %
RedBull	-37,08%	-6,86 %	-18,79 %	-20,62 %	-18,71 %
PopIdol	0,61%	-1,35 %	-2,42 %	-2,28 %	-2,16 %
SkyFade	-12,60%	-2,26 %	-7,80 %	-8,29 %	-7,07 %
SunFade	-23,00%	4,63 %	-5,17 %	-5,01 %	-3,21 %
Tangerine	-10,27%	0,74 %	-5,60 %	-5,37 %	-3,45 %
Average	-6,86%	-1,68%	-4,81%	-5,05%	-4,52%

the TMO and the WP is separable: (i) the TMO optimization, and (ii) the WP optimization. Instead, the temporally optimized TMO requires to jointly optimize the two functions, which requires some simplifying assumptions to compromise between rate and distortion minimization, as well as temporal

regularization. Another benefit of the WP approach is that the SDR can be optimized with respect to some other criterion, not only for compression but rather aesthetic purpose. Therefore, we proposed a new weighted prediction algorithm and new weighted prediction modes to handle more efficiently the

large diversity of brightness variations in video sequences. The proposed algorithm consists in performing a global motion compensation, based on an optical-flow, followed by the computation of different weighted predictions modes or models. The proposed modes consist in computing a global or local non-linear functions to predict the current frame, unlike state-of-the-art methods always based on linear functions. The performances of the proposed algorithm and each mode is evaluated to identify the best ones, then all modes are put in competition in the same encoding process. This competing strategy is made at two levels, frame and CTU, and different modes combinations are studied. The several proposed weighted prediction modes range from a complex (many metadata) but accurate prediction model (local non-linear) to a less complex and rough prediction model (global linear). It has been shown that depending on the image, at least one of the proposed weighted predictor outperforms the WP implemented in HEVC. Therefore, we propose to allow competition between all the proposed WP modes, and this brings BD-rate gain of about 10% with respect to the HEVC reference software and about 3.5% with respect to the HEVC software with existing WP enabled. Indeed, this competition allows to compensate more types of brightness variation. The RD results also show the superiority of non-linear functions, either in global or local weighted prediction, compared to linear ones despite a higher signaling overhead. The RD performances are improved on SDR graded content and also with TMO-generated SDR content, especially when the TMO introduces flickering artifacts. The RD performances were also improved on the HDR level and thus shows that the proposed solution can improve the HDR RD performances while maintaining the SDR temporal consistency.

REFERENCES

- [1] D. Gommelet, A. Roumy, C. Guillemot, M. Ropert, and J. L. Tanou, "Gradient-based tone mapping for rate-distortion optimized backward-compatible high dynamic range compression," *IEEE Transactions on Image Processing*, vol. 26, no. 12, pp. 5936–5949, Dec 2017.
- [2] R. Boitard, D. Thoreau, R. Cozot, and K. Bouatouch, "Impact of temporal coherence-based tone mapping on video compression," in *21st European Signal Processing Conference (EUSIPCO)*, Sept 2013, pp. 1–5.
- [3] Z. Mai, H. Mansour, R. Mantiuk, R. Ward, P. Nasiopoulos, and W. Heidrich, "Optimizing a tone curve for backward compatible high dynamic range image and video compression," *IEEE Trans. on Image Processing*, Jun. 2011.
- [4] A. Koz and F. Dufaux, "Methods for improving the tone mapping for backward compatible high dynamic range image and video coding," *Image Communications*, vol. 29, no. 2, Feb. 2014.
- [5] C. Ozcinar, P. Lauga, G. Valenzise, and F. Dufaux, "Hdr video coding based on a temporally constrained tone mapping operator," in *IEEE Digital Media Industry and Academic Forum*, Santorini, Greece, Jul. 2016.
- [6] S. Pattanaik, J. Tumblin, H. Yee, and D. Greenberg, "Time-dependent visual adaptation for fast realistic image display," in *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH. New York, NY, USA: ACM Press/Addison-Wesley Publishing Co., 2000, pp. 47–54.
- [7] R. Mantiuk, S. Daly, and L. Kerofsky, "Display adaptive tone mapping," *ACM Trans. Graph.*, vol. 27, no. 3, pp. 1–10, Aug. 2008.
- [8] P. Irawan, J. Ferwerda, and S. Marschner, "Perceptually based tone mapping of high dynamic range image streams," in *Proceedings of the Sixteenth Eurographics Conference on Rendering Techniques*, ser. EGSR. Aire-la-Ville, Switzerland, Switzerland: Eurographics Association, 2005, pp. 231–242.
- [9] R. Boitard, R. Cozot, D. Thoreau, and K. Bouatouch, "Zonal brightness coherency for video tone mapping," *Image Communications*, vol. 29, no. 2, pp. 229–246, Feb. 2014.
- [10] V. Sze, M. Budagavi, and G. Sullivan, *High Efficiency Video Coding (HEVC): Algorithms and Architectures*, 1st ed. Springer International Publishing, 2014.
- [11] D. K. Kwon and H. J. Kim, "Region based weighted prediction for real time h.264 encoder," in *2011 IEEE International Conference on Consumer Electronics (ICCE)*, Jan 2011, pp. 47–48.
- [12] S. H. Tsang, T. K. Lee, Y. L. Chan, and W. C. Siu, "Region based weighted prediction algorithm for h.264/avc video coding," in *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2013, pp. 269–272.
- [13] S. H. Tsang, Y. L. Chan, and W. C. Siu, "Region-based weighted prediction for coding video with local brightness variations," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 23, no. 3, pp. 549–561, March 2013.
- [14] N. Rodrigues, V. da Silva, and S. de Faria, "Hierarchical motion compensation with spatial and luminance transformations," in *Proceedings of International Conference on Image Processing*, vol. 3, 2001, pp. 518–521.
- [15] P. Yin, A. Tourapis, and J. Boyce, "Localized weighted prediction for video coding," in *IEEE International Symposium on Circuits and Systems*, vol. 5, May 2005, pp. 4365–4368.
- [16] N. Zhang, Y. Lu, X. Fan, R. Xiong, D. Zhao, and W. Gao, "Enhanced inter prediction with localized weighted prediction in hev," in *Visual Communications and Image Processing (VCIP)*, Dec 2015, pp. 1–4.
- [17] H. Liu, Y. Chen, J. Chen, L. Zhang, and M. Karczewicz, "Local illumination compensation," ITU-T SG16/Q6, Tech. Rep. Doc. VCEG-AZ06, Jun 2015.
- [18] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, May 2011.
- [19] A. Luthra, E. François, P. Lopez, W. Husak, P. Yin, T. Lu, K. Minoo, and A. Tourapis, "Anchors for hdr and wcg tests," ISO/IEC, MPEG, Tech. Rep. N14548, Jul 2014.
- [20] E. François, P. Yin, and J. Sole, "Common test conditions for hdr/wcg video coding experiments," ISO/IEC, MPEG, Tech. Rep. N15793, Oct 2015.
- [21] G. Sullivan, "Common hm test conditions and software reference configurations," JCT-VC, Geneva, Switzerland, Tech. Rep. L1100, April 2013.
- [22] G. Bjontegaard, "Calculation of average psnr differences between rd-curves," 01 2001.
- [23] E. Reinhard, M. Stark, P. Shirley, and J. Ferwerda, "Photographic tone reproduction for digital images," *ACM Trans. Graph.*, 2002.
- [24] J. Ferwerda, S. Pattanaik, P. Shirley, and D. Greenberg, "A model of visual adaptation for realistic image synthesis," in *Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, ser. SIGGRAPH. New York, NY, USA: ACM, 1996, pp. 249–258.
- [25] T. Mertens, J. Kautz, and F. V. Reeth, "Exposure fusion," in *15th Pacific Conference on Computer Graphics and Applications*, Oct 2007, pp. 382–390.
- [26] S. Raman and S. Chaudhuri, "Bilateral filter based compositing for variable exposure photography," in *Eurographics*, Munich, Germany, 2009, pp. 1–4.
- [27] C. Schlick, "Quantization techniques for visualization of high dynamic range pictures," in *Photorealistic Rendering Techniques*. Berlin, Heidelberg: Springer, 1995, pp. 7–20.
- [28] A. Luthra, E. Francois, and W. Husak, "Call for evidence (cfe) for hdr and wcg video coding," ISO/IEC JTC1/SC29/WG11, Geneva, Switzerland, Tech. Rep. N15083, February 2015.
- [29] T. Aydin, R. Mantiuk, and H. Seidel, "Extending quality metrics to full dynamic range images," in *Human Vision and Electronic Imaging XIII*, ser. Proceedings of SPIE, San Jose, USA, January 2008, pp. 6806–6810.
- [30] R. Mantiuk, K. Kim, A. Rempel, and W. Heidrich, "Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions," *ACM Trans. Graph.*, vol. 30, no. 4, pp. 1–14, Jul. 2011.
- [31] Y. Li, D. Min, M. S. Brown, M. N. Do, and J. Lu, "Spm-bp: Speed-up patch match belief propagation for continuous mrfs," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 4006–4014.
- [32] C. Qifeng and K. Vladlen, "Full flow: Optical flow estimation by global optimization over regular grids," *Computing Research Repository (CoRR)*, 2016.



David Gommelet received the Engineering degree from Ecole Supérieure d'Ingénieurs de Rennes (ESIR), France, in 2013 and the PhD degree in Computer Science from Rennes I University in 2018. This thesis took place between INRIA (Institut National de Recherche en Informatique et en Automatique) and Envivio, acquired by Ericsson in 2015 and renamed MediaKind in 2019, in Rennes, France. He is currently a senior engineer at VITEC in Paris, France, working on image and video compression to contribute on future video compression standards.



Julien Le Tanou is currently a Senior engineer within Ericsson's Media Processing Business Line, conducting research into video processing and compression technologies. Since 2012, he has been with Envivio France, which was acquired by Ericsson in 2015, in charge of research and algorithms design of Envivio's real-time video encoding product, with a focus on HEVC video encoder efficiency optimization, UHD and HDR/WCG signal processing. Prior to Envivio, he has been with Orange Labs in France and Dolby Laboratories in USA, working on

new paradigms of image and video compression for contributing to future TV/video standards. He received an Engineering degree in Telecommunication and Multimedia from Institut Supérieure d'Electronique de Paris, France in 2008, and an Advanced Master's degree in Signal, Image Processing, and Pattern Recognition from Telecom ParisTech, France in 2010.



Aline Roumy received the Engineering degree from Ecole Nationale Supérieure de l'Électronique et de ses Applications (ENSEA), Cergy, France in 1996, the Master degree in June 1997 and the Ph.D. degree in September 2000 from the University of Cergy-Pontoise, France. During 2000-2001, she has been the recipient of a French Defense DGA/DRET postdoctoral fellowship and was a research associate at Princeton University, Princeton, NJ. On November 2001, she joined INRIA, Rennes, France. Her current research and study interests include the area

of statistical signal processing, coding theory and information theory.



Michaël Ropert is currently heading the advanced video team at Ericsson in France. He holds a PhD degree from Rennes I University (1995). Previously, (1991-1992) he was teacher in applied mathematics. He has been with France Télécom in the areas of image and video compression for multimedia and digital television (1996-1999). He has been at the initial step of Envivio and leading the video team for several years (2000-2015).



Christine Guillemot is currently Director of Research at INRIA (Institut National de Recherche en Informatique et Automatique) in France. She holds a PhD degree from ENST (Ecole Nationale Supérieure des Telecommunications) Paris (1992). From 1985 to 1997, she has been with France Télécom in the areas of image and video compression for multimedia and digital television. From 1990 to mid 1991, she has worked as "visiting scientist" at Bellcore Bell Communication research) in the USA. Her research interests are signal and image processing, and in

particular 2D and 3D image and video coding, joint source and channel coding for video transmission over the Internet and over wireless networks, and distributed source coding. She has served as Associate Editor for IEEE Trans. on Image Processing (from 2000 to 2003, and from 2014-2016), for IEEE Trans. on Circuits and Systems for Video Technology (from 2004 to 2006), and for IEEE Trans. on Signal Processing (2007-2009). She has served as senior member of the editorial board of the IEEE journal on selected topics in signal processing (2013-2015) and is currently senior area editor of IEEE Trans. on Image Processing. She is IEEE fellow since January 2013.